PB86-217122

Workshop on Model Evaluation Protocols
Chairman's Report

Battelle Pacific Northwest Labs., Richland, WA

Prepared for

Environmental Protection Agency
Research Triangle Park, NC

Jul 86

U.S. Department of Commerce
National Technical Information Service
NTIS

EPA/600/9-86/018
July 1986

PB86-217122

WORKSHOP ON MODEL EVALUATION PROTOCOLS
Chairman's Report

by

W. T. Pennell
Battelle Pacific Northwest Laboratories
Richland, Washington 99352

Interagency Agreement DW89930059

Project Officer

Jack L. Durham
Atmospheric Chemistry and Physics Division
Atmospheric Sciences Research Laboratory
Research Triangle Park, North Carolina 27711

ATMOSPHERIC SCIENCES RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA *(Please read Instructions on the reverse before completing)*		
1. REPORT NO. EPA/600/9-86/018	2.	3. RECIPIENT'S ACCESSION NO. PB86 217122/AS
4. TITLE AND SUBTITLE WORKSHOP ON MODEL EVALUATION PROTOCOLS: Chairman's Report		5. REPORT DATE July 1986 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) W. T. Pennell		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Battelle Pacific Northwest Laboratories Richland, Washington 99352		10. PROGRAM ELEMENT NO. CCVN1A/05-6079 FY-86 11. CONTRACT/GRANT NO. IAG DW89930059
12. SPONSORING AGENCY NAME AND ADDRESS Atmospheric Sciences Research Laboratory - RTP, NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, North Carolina 27711		13. TYPE OF REPORT AND PERIOD COVERED Final 14. SPONSORING AGENCY CODE EPA/600/09
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

This report summarizes the results of an EPA-sponsored workshop that was held to discuss procedures and protocols for evaluating regional-scale acid deposition models. The workshop was the first of three that are planned to assist the U.S. Environmental Protection Agency, the Ontario Ministry of Environment, the Atmospheric Environment Service of Canada, and the Electric Power Research Institute in designing a model evaluation program. The workshop was asked to consider four major topics: (1) procedures to be used in evaluating the performance of acid deposition models and methodologies for applying these procedures; (2) data requirements of these procedures and methodologies; (3) the probable impact of time and budget constraints on the evaluation process; and (4) possible conflicts between client needs and the probable output of the evaluation program.

Each of these topics was considered, and a series of recommendations was made. These recommendations covered the manner in which the evaluation should be conducted, the specific tasks that should be undertaken to effect the evaluation, the data requirements of the recommended evaluation program, and the types of questions that can and cannot be answered by such a program.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED 20. SECURITY CLASS *(This page)* UNCLASSIFIED	21. NO. OF PAGES 27 22. PRICE

EPA Form 2220-1 (9-73)

DISCLAIMER

Although the information in this document has been funded wholly or in part by the United States Environmental Protection Agency under Interagency Agreement DW89930059 to DOE/Battelle Pacific Northwest Laboratories, it does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

ABSTRACT

This report summarizes the results of an EPA-sponsored workshop that was held to discuss procedures and protocols for evaluating regional-scale acid deposition models. The workshop was the first of three that are planned to assist the U.S. Environmental Protection Agency, the Ontario Ministry of Environment, the Atmospheric Environment Service of Canada, and the Electric Power Research Institute in designing a model evaluation program. The workshop was asked to consider four major topics:

- procedures to be used in evaluating the performance of acid deposition models and methodologies for applying these procedures
- data requirements of these procedures and methodologies
- the probable impact of time and budget constraints on the evaluation process
- possible conflicts between client needs and the probable output of the evaluation program.

Each of these topics was considered, and a series of recommendations was made. These recommendations covered the manner in which the evaluation should be conducted, the specific tasks that should be undertaken to effect the evaluation, the data requirements of the recommended evaluation program, and the types of questions that can and cannot be answered by such a program.

ACKNOWLEDGMENTS

This workshop was supported by the U.S. Environmental Protection Agency in cooperation with the Electric Power Research Institute, the Ontario Ministry of Environment, and the Atmospheric Environment Service of Canada.

1.0 Scope and Purpose of the Workshop

The workshop on model evaluation protocols was the first of three planned to assist the U.S. Environmental Protection Agency (EPA), the Ontario Ministry of Environment (OME), the Atmospheric Environment Service (AES), and the Electric Power Research Institute (EPRI) in designing a program to evaluate the performance of regional-scale acid deposition models. The workshop, which was sponsored by EPA, was held in Raleigh, North Carolina on February 11, 12, and 13, 1986. It was attended by a group of 27 scientists representing government laboratories, industrial groups, private contractors, and universities. An attendance list is given in Appendix A.

The workshop was asked to consider four major topics:

- procedures to be used in evaluating the performance of acid deposition models and methodologies for applying these procedures
- data requirements of these procedures and methodologies
- the probable impact of time and budget constraints on the evaluation process
- possible conflicts between client needs and the probable output of the evaluation program.

Each of these topics was considered by the workshop, and a series of recommendations was made. These recommendations covered the manner in which the evaluation should be conducted, the specific tasks that should be undertaken to effect the evaluation, the data requirements of the recommended evaluation program, and the types of questions that can and cannot be answered by such a program. These recommendations are presented in the balance of this report.

In reaching their conclusions, the workshop participants drew on the results of two previous activities: the EPRI-sponsored Utility Deposition Network (UDN) workshop held in November 1985 and the Concept Plan for evaluation of acid deposition models. The Concept Plan resulted from an August 1985 workshop on this subject, which was sponsored by EPA, OME and AES. The UDN workshop was convened to review a technical plan for a deposition monitoring network that would gather the wet deposition and aerometric measurements needed for an operational evaluation of acid deposition models. The UDN workshop made specific recommendations on the type of measurements required and on the methods to be used in making them. These recommendations were incorporated into the recommendations of evaluation protocols workshop.

The Concept Plan, on the other hand, outlined a series of field studies as well as a wet and dry deposition monitoring network that would generate the data needed for both operational

and diagnostic evaluation of the models.[a] This plan, however, did not convert its recommendations into a set of specific tasks nor did it indicate the relative importance of routine monitoring activities, which generate the data for operational evaluation, compared to process studies, which provide the information needed for diagnostic evaluation. The workshop, therefore, spent considerable time in defining these tasks and in discussing the general order of priorities.

2.0 Philosophical Considerations

2.1 CLIENT NEEDS AND CONSTRAINTS

From the clients' point of view, the primary purpose of regional-scale acid deposition models is to provide scientifically defensible tools for analyzing the consequences of alternative strategies for controlling emissions of acid deposition precursors. Reflecting this need, the clients posed a set of four key policy questions questions that they expect the models to be able to address.[b] These questions are:

1. Deposition Loadings: What are the seasonal and annual averages (EPA has suggested an accuracy of ± 50%) of the wet and dry deposition of acidic species (esp., sulfate, nitrate, hydrogen, and ammonium ions) and oxidants to specific states and provinces or to portions of states and provinces?

Note: From a regulatory point of view, emissions and deposition must be resolved for defined geopolitical areas (states and provinces); however, assessment of the effects of acid deposition requires the ability to compute deposition loadings to specific receptor areas. These areas may be considerably smaller than the geopolitical ones.

2. Source Attribution: What is the net (i.e., wet and dry) deposition of sulfur oxides, nitrogen oxides, and VOCs to each receptor area from each geopolitical source area?

Note: The apportionment of responsibility for adverse effects is an important consideration in formulating strategies for emission modifications.

3. Chemical Nonlinearity: How effective are various strategies for controlling sulfur oxides, nitrogen oxides, and VOCs in reducing acid deposition to sensitive receptor areas? For example, is the regional source-receptor response relationship "linear" for sulfur oxides? Do nitrogen oxides and VOC play an important role in governing the chemical response of the atmospheric system?

---

[a] In this report, operational evaluation is defined as the examination of how well a model reproduces actual observations of deposition and concentration fields on the time and space scales needed for assessment work or policy analysis. Diagnostic evaluation is defined as the detailed examination of how well individual components of the model (i.e., modules) simulate actual atmospheric processes.

[b] The relative importance of each question varies from client to client.

Note: The most effective strategy for reducing acidic deposition may not be a linear "roll-back" of sulfur emissions apportioned by the contribution of a given geopolitical area to the total emission of sulfur oxides. For example, if oxidants are important coreactants in acidifying the atmosphere or terrestrial and aquatic systems, controls on the emissions of nitrogen oxides and VOCs may be required. Candidate emission control strategies to be examined as part of the policy formulation process will consist of independent modification of the emissions of sulfur oxides, nitrogen oxides, and VOCs, all as functions of geography and season.

4. Detectability: How great a reduction of source emissions will be required to produce a reduction in acid deposition that can be observed and verified by the clients' monitoring networks?

The primary purpose of the evaluation program, therefore, is to determine whether or not the models are sufficiently accurate to answer these policy questions. The evaluation program must also make this determination within the clients' budget and time constraints.

2.2 OBSERVATIONS AND CONCERNS OF WORKSHOP PARTICIPANTS

During the course of the workshop several issues were raised that could be a source of conflict between the needs of the clients and what the workshop participants think is possible in a realistic model evaluation program. These issues are discussed in the following subsections.

2.2.1 Conflicts between Model Evaluation and Model Development

A major concern of the participants was the conflict between the clients' need for an independent data set to be used in a "hands-off" or "blind" test of model performance and the model developers' need for data to test or develop the models' representations of key physical and chemical processes. This latter need is particularly pressing for the case of regional-scale acid deposition models because some of the key modules in these models are based on hypotheses that have not received adequate experimental testing. Given the time and budget constraints of the federal acid deposition program, it is not possible to satisfy completely the needs of the clients and of the model developers; thus the workshop proposed a possible compromise. A basic reality of this compromise is that any data used in model development is unavailable for performance evaluation.

2.2.2 Problems in Evaluating Complex Models

A problem that troubled several of the workshop participants was separation of the evaluation process into operational and diagnostic phases.[a] It was generally held that all evaluation activities contain a diagnostic kernel: It is not sufficient to determine whether or not model results and observations agree; it is also necessary to determine why.

[a]Workshop participants preferred the terms "integrated" and "mechanistic" to "operational" and "diagnostic"; however, the latter terms will be used in this report for historical reasons.

Consequently, the basic or "operational" evaluation program advocated by the workshop contains a certain diagnostic component.

Another problem faced by the workshop was the entire concept of a diagnostic or component-by-component evaluation of the models. The group felt very strongly that diagnostic experiments having the goal of increasing scientific understanding of critical chemical and physical processes were vital to the development of scientifically credible acid deposition models. However, it was not believed that the type of experiments achievable in a limited one-to-two year program would be able to produce an objective, statistically valid evaluation of individual modules. There are three basic reasons for this conclusion. First, there have been very few good mechanistic studies of transport, transformation, and deposition in the atmosphere; thus, some of the process modules in the models require additional testing before they will be ready for a hands-off evaluation. Second, the evaluation of individual modules requires detailed, intensive observations of the type that can only be generated by special field programs. Such programs are difficult and expensive. The number of meteorological events that could be sampled by such field experiments in a one-to-two year evaluation program would be, of necessity, limited to a small subset of the possible meteorological and chemical conditions important to the acid deposition process. Third, it would be difficult to isolate model errors resulting from errors in input parameters from those resulting from intrinsic errors in the modules, themselves. This difficulty would be especially severe for a short-duration program limited to a small number of independent realizations.

However, the workshop participants agreed that diagnostic experiments would have value in providing a subjective evaluation of the performance of individual model components. If, for example, there were gross errors between the observed behavior of various processes and their representation in a model, they would probably be uncovered by a limited number of diagnostic experiments. Also, if analysis of a succession of diagnostic experiments continued to require substantial modification of a given module in order to achieve acceptable agreement between simulations and observations -- if there were no convergence to a stable model paradigm -- then this lack of convergence would be substantial evidence that the formulation of the module was in error.

2.2.3 Problems in Setting Priorities

One of the charges given the workshop was to set priorities for the model evaluation process; that is, to determine which aspects of the models contained the greatest uncertainty and thus required the greatest emphasis in an evaluation program. The workshop was uncomfortable in setting these priorities because the participants felt that an insufficient number of model sensitivity studies had been conducted to enable an accurate ordering of priorities. As a result, the workshop strongly suggested that sensitivity studies be conducted as soon as possible in order to guide the process of experiment design. Nevertheless, the workshop participants recognized that the clients required preliminary guidance in this area; consequently, the tasks described in the following sections are ordered according to the participants' initial priority judgments.

3.0 General Recommendations

3.1 GOALS OF THE MODEL EVALUATION PROCESS

The workshop participants agreed that the goals of the model evaluation program as stated in Section 2.1 of this report could not be met by a model evaluation program subject to the probable time and budget constraints. Thus, these goals were modified to be more consistent with what is thought to be achievable. These restated goals are given below:

1. Deposition Loadings: Given data from a surface-based monitoring network operating for only a few years and sampling a limited range of chemical conditions, it will not be possible to determine the accuracy to which a given model can predict climatologically valid deposition loadings to a given area.[a] It will, however, be possible to determine whether or not such a model can simulate deposition fields over the period sampled by the network to an accuracy comparable to the uncertainty to which the monitoring network can define the actual deposition field.[b] It should also be possible to quantify, in an objective manner, the level of disagreement between the measurements and alternative predictions: assuming that this disagreement exceeds the measurement uncertainty.

2. Source Attribution: The workshop participants could think of no way of directly testing or evaluating the ability of a model to make this computation. In fact, there was some consternation about the whole idea of source attribution. The models will certainly be able to compute changes in deposition resulting from changes in emissions; and it is possible, by numerically tagging the emissions from a given area, to determine where the emissions from the tagged area may go. However, one must remember that such relationships only apply to a fixed distribution of emissions. Because of the nonlinearity of the chemistry, changing the emissions at any point in the modeled domain may affect source attributions at every other location.

3. Chemical Nonlinearity: There is no way, short of drastically changing existing emission patterns, of truly evaluating the ability of a model to handle this issue. The only solution, therefore, is to perform fairly extensive diagnostic studies that investigate how accurately process modules in the models represent the corresponding natural phenomena. If these modules appear to be fairly accurate, then there will be confidence in the ability of the models to treat issues such as nonlinearity and source attribution.

4. Detectability: There is no way of directly evaluating the ability of a model to make this sort of computation. However, the degree of confidence that one can have in the predictions of the model depends on how well the various process modules simulate the actual atmospheric phenomena.

In summary, an operational model evaluation program will be able to indicate how well the models can simulate current deposition patterns over a time period comparable to the length of the surface monitoring program. A diagnostic evaluation program, on the other hand, will indicate how well the process modules in a given model simulate the important physical and chemical processes in the atmosphere. Diagnostic studies are essential if we

---

[a] The workshop participants assumed that the seasonal and annual averages required by the clients were climatological averages.

[b] This assumes, of course, that some technique exists (such as kriging) that is capable of accurately estimating this measurement uncertainty.

are to have confidence that the models are capable of simulating deposition for emission compositions and distributions significantly different from the current situation

3.2 MANAGEMENT OF THE EVALUATION PROCESS

It is important that the model evaluation process have credibility within the scientific community because this community will ultimately be called upon to embrace or reject the findings of the process. The model evaluation program, therefore, should be managed by a highly qualified, disinterested party. The managing organization should be responsible for developing the evaluation protocols, supervising their execution, and reporting on the results. The National Academy of Sciences was suggested as the most acceptable organization for this role.

3.3 MINIMUM LENGTH OF THE PROGRAM

The workshop participants felt very strongly that the field measurement program associated with the model evaluation process should last a minimum of two years and longer if possible. In the view of the workshop, a shorter measurement program will not gather a sufficient data set for a useful evaluation of the models.[a]

3.4 RELEASE OF THE DATA

The workshop agreed that all data collected during diagnostic studies should be released to potential users as soon as it has been quality audited. None should be sequestered for hands-off evaluation purposes. The reason for this position is the critical lack of experimental data for model development purposes. Filling this need was judged more important than achieving a completely hands-off diagnostic evaluation. The workshop also agreed that the first year of data from the operational evaluation program should be delivered to the users for model development purposes. Only the second year of data should be sequestered for use in blind evaluation tests.

4.0 RECOMMENDED TASKS: FY 1986

The workshop recommended that several preliminary tasks be started immediately because the results are needed for planning the work to be accomplished in FY 1987 and beyond. The recommended tasks are:

[a] For example, experience with current-generation models shows that agreement between model results and observations can vary considerably from year to year. One case was cited in the workshop where a model agreed quite well with one year's data but performed poorly the next year.

4.1 DEVELOP MODEL EVALUATION METHODS

A systematic study is needed of available statistical methods for testing model performance. This study should be subject to strict peer review, because the results will dictate the methods to be used in grading model performance. This study should include an examination of methods for estimating the uncertainties in defining deposition fields using surface network data and of methods for constructing long-term averaged (e.g., seasonal and annual) deposition fields from a limited number of modeling runs.

4.2 DEVELOP EVALUATION PROTOCOLS

A contract should be initiated with the National Academy of Sciences to manage development of the evaluation protocols. This group should also be involved in the peer review of model evaluation methods.

4.3 ANALYZE EXISTING DATA BASES

Existing data bases should be examined for insights into several issues important to model evaluation. These issues include:

- temporal and spatial variability of deposition and concentration fields
- vertical profiles of pollutant species over the modeling domain
- transport of material under various weather conditions.

Possible data bases are OSCAR, AMBIENS, PEPE/NEROS, North Bay, SURE, APIOS, MAP3S, SURE, and CAPTEX. These data bases will require quality auditing before they can be used, however.

4.4 MODEL SENSITIVITY STUDIES

Sensitivity studies are needed in order to determine which components dominate model response and thus require top priority in the model evaluation process.

4.5 EMISSION STUDIES

Several questions were raised about the accuracy of current emission inventories, especially with regards to VOCs, and the cost of achieving a significant improvement in these inventories. Consequently, the workshop recommended immediate work on a modeling study to evaluate the level of agreement between measurements of VOC speciation at rural locations in the eastern United States and predictions, based on existing emission inventories, of simple box models. The workshop also recommended a preliminary study of the uncertainties in existing inventories and the cost of significantly improving them. If, for example, it can be shown that current inventories of certain critical emissions are

uncertain by a significant amount (say a factor of two, or so) then expenditures on reducing these uncertainties would be justified. If, on the other hand, the uncertainties are considerably smaller (e.g., on the order of 20%) then such expenditures are not warranted.

5.0 RECOMMENDED TASKS: OPERATIONAL EVALUATION

The workshop agreed that given current budget restrictions, highest priority should be placed on establishing and operating a surface-based wet deposition and aerometric monitoring network and on performing other tasks required to produce a data base for operational evaluation of regional-scale acid deposition models. The first four tasks are listed in order of decreasing priority. The last three are necessary support tasks for any of the first four. Several of these tasks have a definite diagnostic component, but the workshop participants felt that this component was vital to the interpretation of the operational comparisons.

5.1 DEPLOY AND OPERATE A SURFACE-BASED MONITORING NETWORK

As mentioned above, top priority is given to the task of establishing and operating a 30-station wet deposition and aerometric monitoring network in the northeastern United States. This network should be maintained for a minimum of two years. Recommended measurements for the network are given in Tables 5.1 to 5.3.

Table 5.1
RECOMMENDED PRECIPITATION CHEMISTRY MEASUREMENTS AT STANDARD NETWORK STATIONS

Analyte	Time Resolution	Quantitative Detection Limit (µmol/L)	Desired Accuracy
H^+ (free)	24-hr av	---	±0.05 pH units
Conductivity	24-hr av	---	±20%
Acidity	24-hr av	---	±0.05 pH units
SO_4^{2-}	24-hr av	0.2	±Lgr of QDL or 10%
NO_3^-	24-hr av	0.2	±Lgr of QDL or 10%
Cl^-	24-hr av	2.0	±Lgr of QDL or 10%
NH_4^+	24-hr av	0.5	±Lgr of QDL or 10%
Na^+	24-hr av	0.2	±Lgr of QDL or 10%
K^+	24-hr av	0.2	±Lgr of QDL or 10%
Ca^{2+}	24-hr av	0.2	±Lgr of QDL or 10%
Mg^{2+}	24-hr av	0.2	±Lgr of QDL or 10%
S(IV)	24-hr av	0.2	±Lgr of QDL or 10%

Table 5.2
RECOMMENDED AEROMETRIC MEASUREMENTS AT STANDARD NETWORK STATIONS

Analyte	Time Resolution	Quantitative Detection Limit	Desired Accuracy
O_3	Hourly	5. ppb	±Lgr of QDL or 10%
PAN	Hourly	0.1 ppb	±Lgr of QDL or 20%
NO_2[(a)]	Hourly if chem-illuminescence. 12-hr av if TEA-Impregnated Filter	0.1 ppb	±Lgr of QDL or 20%
Mass, SO_4^{2-}	12-hr av	0.2 ppb (SO_4)	±Lgr of QDL or 20%
Inorganic Nitrate	12-hr av	0.2 ppb	±Lgr of QDL or 20%
HNO_3	12-hr av	0.2 ppb	±Lgr of QDL or 20%
NH_3	12-hr av	0.2 ppb	±Lgr of QDL or 20%
SO_2	12-hr av	0.2 ppb	±Lgr of QDL or 20%

[(a)]At selected stations only.

Table 5.3
RECOMMENDED METEOROLOGICAL MEASUREMENTS AT STANDARD NETWORK STATIONS

Variable	Suggested Accuracy/Precision
Wind speed	±10%/0.5 m/s
Wind direction	±10°/5°
Vertical wind	±0.3 m/s
Temperature	±0.2 C/0.1 C
Pressure	±0.5 mb/0.2 mb
Relative humidity	±2%/1%
Precipitation	±10%/0.01 cm
Ultraviolet radiation	±20%/10%
Total solar flux	±20%/10%
Surface wetness	Not determined

In addition to these measurements, any additional aerometric and meteorological parameters needed to compute dry deposition fluxes will be measured at 5 to 6 specially chosen sites. These will be sites representative of the dry deposition flux over a wide surrounding area.

5.2 VERTICAL PROFILES OVER THE MODELING DOMAIN

In this task, frequent aircraft flights will be made over various portions of the eastern United States. The purpose of these flights is to obtain information on the vertical distribution, from near the surface to several thousand feet, of several important species: SO_2, SO_4^{2-}, H_2O_2, O_3, NO, NO_2, HNO_3, and PAN. In addition to these quantities, measurements of hydrocarbon speciation will be made in the boundary layer if investigation of the PEPE/NEROS data (Section 4.3) show such measurements to be desirable. In order to get a representative distribution of measurements, the flights will be made year-round and under all reasonable meteorological conditions.

5.3 DEPLOY AND OPERATE SUBGRID VARIABILITY NETWORKS

The purpose of this task is to gather data on the subgrid variability of precipitation chemistry and ambient concentration of pollutants. The workshop participants felt that these measurements are very important for interpreting the results of network

measurements. The subgrid variability networks will also form the nucleus for additional diagnostic studies.

The workshop suggested establishment of two subgrid variability networks: one in Kentucky or the Ohio River valley and one on the U. S./Canada border. They should consist of an enhanced central monitoring stations surrounded by a cluster of approximately 100 sequential precipitation chemistry monitoring sites. The cluster would cover an area of 200 by 200 km or about 4 RADM grid cells. The enhanced stations should collect precipitation in 1 mL cuts, and the samples should be analyzed for the species listed in Tables 5.1 and 5.4. The enhanced stations should also make the aerometric measurements listed in Tables 5.2 and 5.5. At the cluster stations, the 1 mL sequential samples will be analyzed for the same analytes as the enhanced stations except organic acids. In addition to precipitation chemistry, integrated gas concentration measurements may be made using inexpensive transition flow samplers should be made at cluster sites. These samplers should also be collocated at the enhanced sites. The full subgrid network should be operated for two 2-month intensive periods per year.[a] Also, about ten of the sites should be operated continuously.

TABLE 5.4
ADDITIONAL PRECIPITATION CHEMISTRY MEASUREMENTS AT ENHANCED NETWORK STATIONS

Analyte	Time Resolution	Quantitative Detection Limit (μmol/L)	Desired Accuracy
Aldehydes	24-hr av	Compound Dependent	Compound Dependent
Organic Acids	24-hr av	Compound Dependent	Compound Dependent
H_2O_2	24-hr av	1	Not Determined

[a]Experience has shown that a two-month field operation would enhance the probability that a sufficient range of meteorological conditions would be sampled by the field measurement program.

Table 5.5
ADDITIONAL AEROMETRIC MEASUREMENTS
AT ENHANCED NETWORK STATIONS

Analyte	Time Resolution	Quantitative Detection Limit	Desired Accuracy
Formaldehyde	3-hr av	0.3 ppb	±Lgr of QDL or 20%
Reactive HC	3-hr av	0.1 ppb	±Lgr of QDL or 20%
H_2O_2	1-hr av	0.1 ppb	±Lgr of QDL or 20%
Size segregated SO_4^{2-}	3-hr av	0.1 mg/M^3	±Lgr of QDL or 20%

During the intensive observation periods, aircraft measurements should be made of the vertical profiles of key chemical species. These should be the same species as measured at the enhanced sites. Additional measurements of the vertical structure of SO_2 and O_3 might be made by an aircraft equipped with a downward-looking uv DIAL, if this technique proves to be economically attractive.

5.4 EMISSION INVENTORIES

In this task existing emission inventories should be updated to correspond to the time periods being modeled in the operational evaluation studies. In addition, more extensive improvements should be made to the inventories, such as improving VOC inventories, if the FY 1986 emission inventory task shows that this effort is justifiable economically.

5.5 SUPPORT TASKS

In addition to these major tasks there are three additional support tasks. These tasks are:

- Quality Audit Data
- Archive Data
- Perform Evaluations.

The scope of each task is obvious from its title.

5.6 SUGGESTED PROTOCOLS

As indicated throughout the previous discussion, the workshop made several suggestions about the evaluation process and the distribution of data: For the first year of

measurements, all data should be released as soon as quality auditing is complete. In the second year, the following policies are suggested:

1. Emission inventory updates quality audited and released to modelers

2. Surface monitoring data quality audited and sequestered

3. Subgrid variability data quality audited and released to modelers

4. Vertical profile data (over domain and over subgrid network) quality audited and sequestered

5. Comparisons of model outputs and data overseen by the National Academy of Sciences
   - surface comparisons made by approved objective techniques
   - vertical profile comparisons made by objective and subjective analysis

6.0 RECOMMENDED TASKS: DIAGNOSTIC EVALUATION

The primary purpose of the diagnostic evaluation process is to assure that the models are providing an accurate simulation of the physical and chemical process that control the transport, transformation, and deposition of acidic materials. Initially, seven types of diagnostic studies were identified as necessary for assessing the ability of acid deposition models to simulate wet and dry removal in a realistic manner. Two of these study classes, subgrid variability and vertical profiles, were included in the basic operational evaluation program. The remaining five study classes were grouped according to a preliminary assessment of their relative importance. The highest priority was given to studies of the wet and dry deposition modules. A lower priority was given to studies of transport, gas-phase chemistry, and treatment of the inflow boundary conditions. Each of these study classes will now be discussed in more detail, and the reasons for their priority ranking will be given.

The diagnostic studies identified during the workshop represent work that needs to be done in addition to the Department of Energy's PRECP program. The PRECP program is considered an important source of mechanistic information needed by model developers; however, the project is not a diagnostic model evaluation program.

6.1 HIGH PRIORITY STUDIES

6.1.1 Wet Deposition Module

These studies are needed to ensure that the physical and chemical processes governing the wet removal of acidifying gases and aerosols are accurately represented in the models. To accomplish this evaluation, intensive precipitation scavenging studies should be conducted

in the two subgrid regions during the previously mentioned intensive observation periods. The necessary data base will be provided by the subgrid surface network and aerometric aircraft. Additional measurements would be required from radar (to map cloud and precipitation fields), cloud-physics equipped aircraft, and enhanced upper-air soundings.

Analysis of the data can proceed by simulating the observed cloud and precipitation chemistry fields with diagnostic models. Comparing observed and simulated data will give an indication of how well the model is representing the critical processes. Alternative parameterization schemes can be tried to determine if they produce superior simulations. Given a sufficient number of case studies, those parameterization schemes that consistently result in the best representation of the observations should emerge.

6.1.2 Dry Deposition Module

Representation of dry deposition by the models is probably the most uncertain of all of the process modules. However, it is also the most difficult to measure, especially on the regional scale. The strategy for evaluating this module would be to compare model simulations with alternative methods for measuring or estimating dry deposition in the atmosphere. In making these comparisons, the dry deposition core stations should be used as the primary diagnostic reference points. The NAPAP dry sites can supplement these stations as a secondary reference. Additional core stations should be placed at each of the two enhanced subgrid sites, and additional measurements for deducing dry deposition from air concentration and meteorological measurements should be made at, say, four of the subgrid cluster stations.

Using combinations of ground-based and aircraft data, estimates can be made of dry deposition fluxes to each of the subgrid areas. These estimates can then be compared with model computations. Similar comparisons can be made at each of the core sites and at the NAPAP dry sites.

6.2 LOWER PRIORITY STUDIES

6.2.1 Atmospheric Transport

The ability of the models to handle transport, especially through storms, is a primary concern of the model evaluation process. The principal reason that this class of studies is given a lower priority than the previous two is that past and currently planned studies such as CAPTEX and ANATEX may provide a sufficient data base for evaluating the ability of the models to treat this process. In evaluating transport, two types of studies are envisioned: studies of long-range horizontal transport and studies of vertical translation in storms.

The most effective tests of a model's ability to handle pollutant transport are tracer studies. In these tests, a point-source "tracer" is added to the list of dependent variables simulated by the model. The model is then used to simulate the transport of the tracer material. The

first tests of the regional-scale acid deposition models should be against the CAPTEX results (this testing should get under way in 1986). The next step in the evaluation process should be to examine the ANATEX data. If these data are suitable, a similar set of comparisons should be made. If the ANATEX data are not suitable, a special tracer release program will be required. In this program, an inert tracer should be released from an appropriate location in the modeling domain for a range of meteorological conditions. The tracer should then be tracked using surface and airborne sampling systems, as was done in CAPTEX.

In analyzing tracer experiments, model-predicted and measured tracer distributions are compared. If the model is not capable of simulating transport of simple inert tracers, it will certainly not be capable of handling the transport of reactive, depositing substances. If, however, the simulations and measurement are in reasonable agreement the confidence in the model's ability to simulate long-range transport will be greatly enhanced.

Studies of through-storm transport can be accomplished by tracking the transport of either an intentionally released tracer material or a conservative tracer-of-opportunity through a number of storm systems. Investigations of both convective and frontal storms should be attempted. Although studies of this type have already been conducted as part of the PRECP program, additional through-storm tracer experiments are needed.

Analysis of the data from through-storm transport experiments is accomplished in a similar manner to other tracer studies. That is, transport through the storms is simulated numerically and compared to the observations. Collection of the necessary data, however, could be difficult, particularly for frontal storms. The sampling system must be composed of several tracer-sampling aircraft in addition to a ground-based sampling network. At least one of these aircraft will have to be capable of reaching the altitudes of storm outflow.

6.2.2 Gas Phase Chemistry

Aside from the previously-noted information gaps in hydrocarbon emission inventories, there are two major sources of uncertainty in gaseous-phase chemistry submodels:

- the absence of information on reaction-rate parameters and mechanisms of formation for chemical species for which rate measurements have not been performed, and

- probable errors in information corresponding to species for which measurements are available.

The first of these sources of error is particularly pronounced for some of the moderately reactive hydrocarbons, particularly the aromatics, which have received secondary emphasis in previous smog-chamber investigations. Because the more reactive species, such as the olefins, are likely to become chemically exhausted over larger time and distance scales, these moderately reactive compounds are expected to be relatively important participants in regional-scale chemistry.

The second source of error derives from a number of issues. One of the most important of these is the concentration ranges at which laboratory measurements are usually made. As a result, even though smog-chamber and more refined kinetic measurements are available for many of the reactions of interest, these data pertain largely to pollutant-concentration ranges exceeding atmospheric values by two orders of magnitude or more. Moreover, experimental errors such as reactor wall-effects contribute substantially to uncertainties in a number of cases. Finally, mechanistic uncertainties remain which have continued to defy measurements and analysis. An important example of this is our present inability to close atmospheric nitrogen-oxide material balances in a satisfactory fashion, suggesting that important NO_y species may exist, which remain to be identified.

The diagnostic evaluation suggested for the present effort is largely indirect, and involves detailed measurement of hydrocarbon species, reaction products and NO_x/NO_y chemistry at the central stations. These detailed measurements will be compared with zero-dimensional reaction-chemistry simulations, using the parameterizations employed by the models, as well as more elaborate descriptions. Comparisons of key ratios of reactants and intermediates will be employed as the primary tests of reality in the submodel calculations.

An alternative experimental approach is to perform Lagrangian-type experiments in which an air parcel is marked by a tetroon or smart balloon. The chemical evolution of this air mass is then simulated by the chemistry module and compared with observations.

6.2.3 Boundary Conditions

The purpose of these studies is to determine whether or not model simulations are being affected by the transport of errors into the modeling domain through the inflow boundaries. A possible method of attacking this problem would be to make a series of research aircraft flights along the inflow boundary at various altitudes in the boundary layer.[a] These flight tracks would be interrupted occasionally by spiral soundings at pre-selected points. The measurements made by the aircraft would include the usual suite of aerometric measurements. Analysis of the results would proceed by comparing the observed and synthesized boundary conditions and by comparing model results from simulations driven by these different boundary conditions.

[a] There was considerable discussion about the practacality of such flights, however.

APPENDIX A

List of Attendees: Model Protocols Workshop
February 11-13, 1986
Raleigh, N.C.

Chandrakant Bhumralkar
Sr. Scientist
R/PDC NOAA
6010 Executive Blvd.
Rockville, MD 20852
301-443-8108

Greg Carmichael
Chemical & Materials Engr. Program
129 C.B.
University of Iowa
Iowa City, IA 52242
319-353-4193

Julius Chang
National Center for Atmospheric Research
Boulder, CO 80307
303-442-2210

Jason Ching
Meteorology Research MD-80
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711
919-541-4801

Al Christie
Atmospheric Environment Service
4905 Dufferin Street
Downsview, Ontario
Canada M3H 5T4
416-667-4981

John Clarke
Meteorology Research MD-80
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711
919-541-3660
FTS:629-3660

Dave Davies
135 Aspen
Dollard des Ormeaux, Quebec
Canada H9A 2N9
514/683-7952

Ken Demerjian
Atmospheric Sciences Research Center
State U of NY
100 Fuller Rd.
Albany, NY 12222
518-442-3820

Robin Dennis
Meteorology Research MD-80
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711
919-541-2870
FTS:629-2870

Jack Durham
Atmospheric Sciences MD-57
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711
919-541-2183
FTS: 629-2183

Jeremy M. Hales
Battelle, Pacific Northwest Laboratories
PO Box 999
Richland, WA 99352
509-376-8447
FTS 444-8447

Bruce B. Hicks
NOAA - Air Resourses Laboratory
Atmos. Turbulence & Diffusion Div.
456 S. Illinois Ave.
P.O. Box E
Oak Ridge, TN 37830
615-576-1232

Cynthia Hirtzel
Dept. of Chemical & Environmental Engr.
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
518-266-6191

Warren Johnson
Director, Atmos. Science Ctr.
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
415-859-4755

John A. Laurmann
Gas Research Institute
8600 W. Bryn Mawr
Chicago, IL 60631
312-399-8299

Mei-Kao Liu
Systems Applications, Inc.
101 Lucas Valley Road
San Rafael, CA 94903
415-472-4011

Gregory J. McRae
Chemical Eng.
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213
412-268-2254

P. K. Misra
OME
125 Resources Road
P.O. Box 213
Rexdale, Ontario
Canada M9W 5L1
416-248-3245

Tony Olsen
Battelle, Pacific Northwest Laboratories
PO Box 999
Richland, WA 99352
509-376-4265
FTS 444-4265

Marvin Olson
Air Quality Environmental Research Branch
Atmospheric Environment Service
4905 Dufferin Street
Downsview, Ontario
Canada M3H 574
416-667-4903

William Pennell
Battelle, Pacific Northwest Laboratories
PO Box 999
Richland, WA 99352
509-376-8330
FTS 444-4265

Dave Renne
Battelle, Pacific Northwest Laboratories
PO Box 999
Richland, WA 99352
509-376-5063
FTS 444-4265

Steve Seilkop
Analytical Sciences, Inc.
P.O. Box 12794
Research Triangle Park, NC 27709
919-549-8330

Akula Venkatram
ERT Inc.
975 Business Center Circle
Newbury Park, CA 91320
805-499-1922

Eva Voldner
Air Quality Environmental Research Branch
Atmospheric Environment Service
4905 Dufferin Street
Downsview, Ontario
Canada M3H 574
416-667-4788

Boris Weisman
M.E.P. Co.
7050 Woodbine Ave. Suite 100
Markham, Ontario
Canada L3R 4GH
416-477-0870

Paolo Zannetti
Aero Vironment Inc.
Manager, Environmental Analysis
and Software Dept.
825 Myrtle Avenue
Monrovia, CA 91016
818-357-9983